様式 1-4-2 (2022-1)

令和５年度研究開発成果概要書

採 択 番 号　　22901
研究開発課題名　プライバシー保護連合学習の高度化に関する研究開発
副　　題　　継続実運用に資する不正取引モニタリングに向けたプライバシー保護連合学習の高度化

（1）研究開発の目的

複数の銀行から提供される顧客データから求められる特徴量と犯罪フラグの共通化を行い、参加金融機関数に増減があっても、安定した継続学習を行える学習アルゴリズムの開発とそのロバスト性向上のための銀行取引履歴と口座情報の疑似データ生成アルゴリズムを提案する。さらに、不正送金検知における AI の回避攻撃とデータ汚染攻撃のシナリオを数種類想定し、その対策に関する先行研究を調査する。

（2）研究開発期間

令和４年度から令和６年度（３年間）

（3）受託者

国立大学法人神戸大学＜代表研究者＞
ＥＡＧＬＹＳ株式会社

（4）研究開発予算（契約額）

令和４年度から令和６年度までの総額 56 百万円（令和５年度 16 百万円）

（5）研究開発項目と担当

研究開発項目１　DeepProtect の高度化に関する研究
1-1　実運用を模擬したテスト環境の開発と性能評価（神戸大学）
1-2　DeepProtect の継続学習化（ＥＡＧＬＹＳ）
1-3　DeepProtect の敵対的サンプルとデータ汚染攻撃への耐性向上（神戸大学）

研究開発項目２　DeepProtect を用いた不正取引検知エンジンの開発
2-1　データパイプライン構築モジュールの開発（ＥＡＧＬＹＳ）
2-2　DeepProtect 学習実施モジュールの開発（ＥＡＧＬＹＳ）

研究開発項目３　継続実運用を想定した不正送金検知実証実験環境の整備
3-1　不正送金検知特徴量と不正判定基準の標準化（神戸大学）
3-2　人間系フィードバックを容易とする支援ツールの開発（ＥＡＧＬＹＳ）

（6）特許出願、外部発表等

		累計（件）	当該年度（件）
特許出願	国内出願	0	0
	外国出願	0	0
外部発表等	研究論文	0	0
	その他研究発表	3	2
	標準化提案・採択	0	0
	プレスリリース・報道	0	0
	展示会	0	0
	受賞・表彰	0	0

（7）具体的な実施内容と成果

研究開発項目１　DeepProtect の高度化に関する研究

1-1．令和４年度に引き続き、DeepProtect の高度化を効率的に行うテスト環境を構築するため、口座取引を模擬する疑似データの生成方法を検討した。本年度では、口座取引を模擬する時系列生成モデルを定式化し、モデルパラメータと各口座の潜在変数（入出金特性）を推定して疑似データを生成する機械学習手法を開発した。本年度では、ベンチマーク用の口座取引データで動作検証を行い、多様な悪性取引の表現を学習できる可能性があることを検証した。次年度において、銀行から提供された実データを使ったモデル学習および疑似データ生成を行い、提案手法を性能評価し、実データに近い疑似データが得られるよう改善を行う。

1-2. DeepProtect を使ったシステムの長期運用を見据え、令和４年度に開発した破壊的忘却を回避する継続学習と連合学習を同時に達成する学習アルゴリズムの詳細な検証として、継続学習に関する問題について、精度検証を実施し、対外発表のための準備を実施した。また、このアルゴリズムを銀行取引データに適用するための前準備を実施した。前準備として、データ提供された１銀行の取引データについて、特徴量の精査、分析モデルの構築・検証を実施した。次年度においては、追加される他行のデータ解析を進め、継続学習の実施に関する検証を行う。

1-3. 不正送金検知連合学習 AI に対し、不正・不審取引を正常取引に誤判定させる回避攻撃が可能かをベンチマークデータにより検証した。具体的には、連合学習に参加する一組織がデータ汚染攻撃を受け、特定顧客の口座取引データが意図的に操作されることで、連合学習 AI のグローバルモデルの検知を無効化する敵対的サンプルを Sign-OPT アルゴリズムで生成した。次年度において、銀行から提供される実データを使った DeepProtect の脆弱性を評価し、その対策を行う。

研究開発項目２　DeepProtect を用いた不正取引検知エンジンの開発

2-1.DeepProtect を銀行で運用するために、様々な銀行の環境に対応可能なようプロトタイプシステムのチューニングを実施した。令和４年度に開発実施したプロトタイプは、ドッカー上にシステムをホストするシステムであり、システムのホスト環境によっては、銀行運用担当者が計算機の管理者権限を有する必要があり、計算機の管理者と運用者が異なる場合に対応が困難であった。そこで、これまでに開発した特徴量抽出を実施する ETL モジュールのネイティブアプリケーション化を進め、種々の運用体制に対応可能となるように改良を実施した。

2-2. Deep Protect では各ローカルモジュールでの学習が参加銀行ごとに同期される必要があり、学習の進捗は最も計算能力が低い参加者に合わせる必要がある。参加銀行間で学習の実施能力差が著しい場合でも、ローカル環境に用意される種々の計算機に対応し、素早く学習が進むように令和４年度に開発した学習モジュールを NVIDIA 社製の CUDA 環境に依存しないように改良を実施した。これにより、GPU を有さない銀行が参加した場合でも、その銀行の学習が素早く実施できるようになり、システム導入の際に銀行からの幅広い要求に対応可能となることが考えられる。

研究開発項目３　継続実運用を想定した不正送金検知実証実験環境の整備

3-1　高精度な不正送金検知を行うため、顧客口座と口座取引の特徴量、さらに不正判定基準の標準化を試みる予定であったが、データ提供が遅れているため、本項目は実施せず、次年度において実施することとした。

3-2. 参加銀行にプロトタイプシステムの説明およびヒアリングが実施できるよう、説明用動画の撮影、ヒアリング項目の取りまとめを実施した。また、開発項目2-1と同様に令和５年度に開発した人間系のフィードバックの取り込みを容易とする支援ツールについて、ネイティブアプリケーション化を実施し、種々の運用に対応できるように改良を行なった。

（8）今後の研究開発計画

令和 6 年度では、実運用に資する DeepProtect の継続学習方式を実現し、複数の銀行から提供されるデータを用いて、データ提供期間における不正検知の再現率が 90%以上維持されることを実証する。また、銀行が提供するデータ項目から求められる特徴量と犯罪フラグの共通化を可能な限り行い、データ項目が完全に一定していなくても、疑似データ生成による欠損補完を行い、安定した継続学習が実現可能であることを実証する。さらに、不正送金検知における AI の回避攻撃とデータ汚染攻撃のシナリオを数種類想定し、その対策を施しても、なお不正検知の再現率 90%の目標が達成できるかを検証する。具体的な実施項目を以下に示す。

研究開発項目 1　DeepProtect の高度化に関する研究

・研究開発項目 1-1　実運用を模擬したテスト環境の開発と性能評価（4月～9 月）

令和 5 年度に開発した状態空間モデルを使った疑似データ生成の性能評価を銀行から提供された実データで行い、実データに近い疑似データが得られるよう改善を行う。これにより、銀行不正送金検知の実運用を模擬したテスト環境を完成する。

・研究開発項目 1-2　DeepProtect の継続学習化（4月～9 月）

令和 5 年度に提供された銀行データに加え、追加される他行のデータ解析を進め、開発アルゴリズムに関する継続学習の検証を行い、銀行取引データに関する連合学習システムの継続学習化の検証を行う。

・研究開発項目 1-3　DeepProtect の敵対的サンプルとデータ汚染攻撃への耐性向上（4月～9 月）

連合学習 AI に対して、不正・不審取引の検知回避につながる敵対的サンプル生成のアルゴリズムを完成し、その攻撃に対する耐性を有する DeepProtect の継続学習アルゴリズムを完成する。また、銀行から提供された実データを使った性能評価実験を実施し、攻撃への耐性能力を定量的に評価する。

研究開発項目２　DeepProtect を用いた不正取引検知エンジンの開発

・研究開発項目 2-1　データパイプライン構築モジュールの開発（4月～9 月）

開発したモジュールの導入に関するヒアリングを実施し、参加銀行の環境に合わせたシステムチューニングを実施する。

・研究開発項目 2-2　DeepProtect 学習実施モジュールの開発（4月～9 月）

開発したモジュールの導入に関するヒアリングを実施し、参加銀行の環境に合わせたシステムチューニングを実施する。

研究開発項目３　継続実運用を想定した不正送金検知実証実験環境の整備

・研究開発項目 3-1　不正送金検知特徴量と不正判定基準の標準化（4月～9 月）

参加金融機関が提供するデータ属性値の標準化と不正判定基準の標準化を行う。

・研究開発項目 3-2　人間系フィードバックを容易とする支援ツールの開発（4月～9 月）

開発した支援ツールに関して、参加銀行に対するヒアリングを実施。ヒアリングに基づくフィードバック支援ツールの改良を行う。